DOCUMENT RESUME

ED 074 072 TM 002 431

AUTHOR Bruno, Nancy L.; And Others
TITLE Statewide Assessment: Methods and Concerns.
INSTITUTION Educational Testing Service, Princeton, N.J. Center for Statewide Educational Assessment.
PUB DATE 72
NOTE 29p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Data Collection; Educational Improvement; *Educational Status Comparison; *Evaluation Methods; Guides; *State Surveys; *Student Evaluation

ABSTRACT

Some suggestions and ideas for initiating a statewide educational assessment program are provided. Guiding principles for an assessment should be: (1) to specify and define educational goals in terms of measurable outcomes; (2) to involve various publics extensively; (3) to use measurement instruments having face and content validity; (4) to include noncognitive student behaviors; (5) to present the results in a form understandable by those outside the professional education community; and (6) to view assessment not as an end but as a means of providing useful information to decision makers. The objectives of the initial statewide assessment program should be: (1) to collect student performance data that can provide a status report on the quality of education in those goal areas identified as having high priority; (2) to introduce the concept of assessment and its usefulness as a source of information for both decision makers and concerned parents or taxpayers; (3) to provide a starting point whereby those managing the statewide effort may gain useful experience in operating the program; and (4) to develop a method of data analysis that can illustrate the variability of performance due to individual differences among students and to the social context in which they live. Essential components of the assessment should include the provision for an advisory committee and for selecting its members, conducting meetings to inform citizens and school personnel of the nature, scope, and methodology of the program, data collection materials, data analysis procedures, and reporting strategies. (Author/DB)

ED 074072

CSEA

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
NANCY
BRUNO
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE U.S. OFFICE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Statewide Assessment: Methods and Concerns

Nancy L. Bruno
Paul B. Campbell
William H. Schabacker

TM 002 431

CENTER FOR STATEWIDE EDUCATIONAL ASSESSMENT
EDUCATIONAL TESTING SERVICE • PRINCETON, NEW JERSEY

Copyright © 1972 by Educational Testing Service
Educational Testing Service is an Equal Opportunity Employer.

STATEWIDE ASSESSMENT:

METHODS AND CONCERNS

Nancy L. Bruno
Paul B. Campbell
William H. Schabacker

TABLE OF CONTENTS

	PAGE
Introduction	1
Assessment Principles	1
A Model for Beginning State Assessment	4
Components	6
Data Collection	8
Analysis Procedures	10
Reporting Results	12
Inter-District Comparisons	14
Sample Design and Data Analysis	14
Criticality of Condition Variables	15
Summary	17
References	19
APPENDIXES	
A: Sample Selection - Individual Student Sampling Unit - Equal Probability	20
B: Options for Reporting Results	21

INTRODUCTION

Traditionally the method of assessing the quality of educational programs and services, a long standing interest of state government, has been to examine such things as the quality of buildings and facilities, the credentials of professional personnel and per pupil expenditures.

The passage of the Elementary and Secondary Education Act of 1965 with its requirement that schools objectively assess the effects on student achievement produced by federally funded programs for the educationally deprived, focused attention on measuring the performance of students to assess the effectiveness of the schools.

The essential merit of this approach has become increasingly evident to educational decision makers at the state level, and laws mandating statewide assessment of the quality of education have been passed in many states.

Due to time constraints, inadequate budgets, and the speed with which many assessments have been mandated, state departments of education often find they lack a workable plan for assessment or the personnel to conduct it.

ASSESSMENT PRINCIPLES

Experience with statewide assessment programs has led us to the formulation of the following principles which are designed as a guide for state department personnel and others to assist them in optimizing the chances for a successful assessment.

Involve the Community. Effective educational assessment demands the recognition and involvement of the entire community; legislators, educators, parents, students, business managers, labor leaders and other concerned groups.

One method of involving them is to have representatives from each group assist in determining what the goals for education ought to be. Since each group may have different priorities this could be a time consuming activity. The time will be well spent, however, since in addition to determining the goals, the participants should also become aware of the needs and constraints of the others. For example, the legislator wants to know about how much pupil learning and development the money he appropiates for education is buying. He also must answer to his constituents who may not reelect him if they feel he is not concerned about the quality of the education their children are getting. Since the goals for education most directly concern the students, they should have representation in deciding what those goals ought to be. Parents want assurance that their children are receiving the kind of education that will enable them to cope with the ever increasing complexity of the world in which they live.

Teachers also have an interest in assessment. Some may have negative attitudes because they feel they personally will be evaluated. In addition to the valuable contribution they can make, they will be less apt to feel

threatened because they have been given the opportunity to participate in the developmental phases of the program.

The early involvement of the various interest groups should facilitate understanding and cooperation when the assessment is conducted.

Specify and Define Goals. After the goals have been determined, they must be defined operationally and behaviorally so they can be measured. The community should continue to be consulted in this phase; especially the educators.

An example of this type of definition is the goal "To appreciate human endeavor in the arts." One aspect of this goal would be to appreciate music. An appreciation of music could be defined in behavioral terms as the number of times tapes and records are used. This definition corresponds to the receiving and responding levels of the affective domain (Krathwohl, et al, 1964). The behavioral objective could then be measured by a frequency count of the tapes and records used in the library and those taken out for off campus listening. The number of usages and the proportion of students involved would be an indicator of the student body's appreciation of music.

Measuring devices must have face and content validity. The instruments should contain an adequate sampling of the specified universe of content. In addition, they should be face valid. That is, the layman must be able to look at the tests and see the relationships between them and the goals being measured. If the objective is to measure understanding and the instrument contains items that are purely factual in content, the instrument would not have content validity although it might appear to be face valid. Adequate assessment devices must present both.

Take noncognitive effects of school into account. Society is delegating more and more responsibility to the schools for developing learning outcomes

which are not skills centered. The appreciation of music goal mentioned earlier is one example. Another is the development of a positive self-concept. Although these noncognitive areas are admittedly more difficult to measure, in an assessment program they must not be ignored in the early phase or they most likely will continue to be neglected as the program is enlarged.

Data presentations should be designed for lay understanding. Possibly the most crucial aspect of a successful assessment program is the reporting of results. The reports should be in terms that are understandable to the layman. Interpretation of statistical data, particularly that which requires qualification, such as test scores, is most effective when interaction between the receiver and the presenter is possible. However, there is likely to be little interaction if the results are reported in sophisticated technical terms. Four possible alternatives for use in the presentation of data are: expectancy tables based on previous year's performance; comparison with state norms; percentage of response to each option of key items; and description of the distribution of student scores in terms of the kinds of problems they are solving successfully and the kinds which are presenting difficulty.

Assessment must not be an end in itself. The last principle, which perhaps should have been first, is that assessment must be clearly identified as one component of the total education process. Evaluative data are collected to meet specific needs and if the data are not related to these purposes they are useless. Assessment must provide feedback to enable decision makers at various levels to make program modifications necessary for educational improvement.

A MODEL FOR BEGINNING STATE ASSESSMENT

The politics of assessment frequently limit the number of methods available to achieve the principles which were set forth in the preceding section. As a result, state department personnel may have to work under any or all of the following constraints.

First, time schedules -- especially when limited by legislative action -- most often do not allow an adequate and thorough development of assessment procedures.

Second, the resources made available are usually far less than required. Thus, one must expect that compromises will be made in the operations of the program.

Third, the unavailability of adequate professional staff further complicates effective implementation of assessment activities.

Fourth, the conceptions of assessment held by the several publics who are concerned with it are frequently ambiguous and overly optimistic.

Within the context of the constraints identified a simplified model with limited objectives can be implemented. The first objective is to collect data which will provide a status report on education within the state in specified areas of greatest interest. This model is designed to provide statewide data not individual or school building data. The most commonly specified areas are reading and math because these basic skills are fundamental to most educational activities. It is recommended, however, that even the first level model include data collection in at least one noncognitive area. This recommendation is made because of the human tendency to concentrate efforts upon the areas being evaluated. Therefore, the failure to evaluate noncognitive areas has the effect of

focusing the educational process on the skill development segment of education to the neglect of the equally important but more difficult to measure noncognitive areas.

One such area related to the two major cognitive areas, is student attitude toward school or learning. This attitude has apparent value to those concerned with the educational process and seems a reasonable area of interest in which to begin initial efforts in noncognitive data collection.

The second objective of this first level model is to introduce operational concepts of assessment to the interested parties. These include school personnel, both administrative and teaching, state government personnel, including legislators and executive office staff, and concerned community groups. The consideration of methods of statewide educational data collection by these groups will provide them with insight into the limited nature of the kind of information available from a basic model and the problems related to obtaining it.

The third objective is to provide a plan which will enable state education personnel to gain experience in dealing with assessment problems. Among these problems are: selecting and securing data collection devices; negotiating the needs for assessment data presented by a heterogeneous public; developing communication strategies which minimize destructive conflict and optimize data utilization; and establishing the organizational structure to carry out assessment programs.

The fourth objective is to provide a method of analyzing the data to illustrate the variability of performance due to the individual differences among students and to the social context in which they live. The reason for using this method of analysis is to clarify ambiguous perceptions of the interested publics concerning these correlates of performance. Most people recognize that

there are individual differences and that these differences contribute to the difficulty of learning. On the other hand much of the public appears to expect that some simple method of instruction, if properly applied, will overcome such difficulties. The data produced by this model can and should be used to analyze contextual learning difficulties and differences, thereby requiring a major consideration of them.

Components

The components of the first level assessment model include introductory activities, data collection materials, analysis procedures, and reporting strategies.

Introductory activities are informative and communicative in purpose. An integral part of these activities is the selection of an Advisory Committee to consider the objectives of the assessment and plans for its accomplishment.

The selection of the Advisory Committee should be given very careful thought. It should be cross-sectional in nature, representing the several publics who are concerned with assessment. An additional qualification for members should be sincere interest in education as a community responsibility. It is probably not possible under the conditions for which this model was designed, to convene the Advisory Committee for the several sessions which would result in optimum consideration and support of the assessment program. Therefore, the committee function would be that of a review group. Department staff, augmented as necessary by independent consultants, should prepare tentative objectives and plans for committee review. The committee's recommendations should be carefully considered, incorporated if possible, and always given the courtesy of a response.

Following the preparation of initial plans and review by the Advisory Committee, a series of regional conferences should be conducted. The conferees should represent both the community and professional educators. Special

emphasis should be placed upon the effects of assessment activities in the school, and open discussion of this matter should be given substantial time on the conference agenda.

The arrangements for each conference should include provision for small group sessions to facilitate an interchange of ideas among the representatives of the different groups. The care with which the Advisory Committee was selected will be reflected in the success of these conferences because, to the degree that the participants feel they were represented in the earlier planning, they will be inclined to respond favorably to the plans.

The success of these introductory activities also depends on the content of the plans presented. The components which delineate the desired content are described next.

The major component is the measurement package. For the objective of this first level model, presently existing devices should be selected where possible. Standardized achievement tests or minor modifications of such tests are appropriate for reading and math. The exercises used by national assessment, which are placed in the public domain upon release, are also candidates for utilization. For example, reading and math exercises are available in standardized tests, released NAEP items and the Delaware and Michigan state assessment programs. The Delaware, Michigan and Pennsylvania assessment programs have also used instruments that measure attitude toward learning.

In order to clearly communicate what will be accomplished with the measurement package, it is advisable to provide a content reference which will show how scores may be reported in specific skill or behavior related form. Content reference, as used here, means an example or description of the behavior implied by the response to the question. For example, the steps to provide such a reference for a sample reading comprehension item based upon paragraphs are as follows:

.Identify a series of skills tested by the specific questions asked, such as making inferences, detecting mood, or recognizing factual detail.

.Prepare or select an equivalent paragraph to that presented in the test and show how it will be used to illustrate the obtained results.

.After testing, the percentage of students who successfully respond to each kind of question, indicating accomplishment of the skill of interest should be reported.

In the case of attitudinal questions, descriptions of specific approach or avoidance behaviors, evoked by situations or persons, are the content reference. A Likert type response format to this sort of item will provide information about the proportion of students who respond in each direction (approach or avoid) and also about the intensity of the responses to the behavior in question. After testing, if it is decided not to report the actual questions, the underlying constructs can be described and the response frequencies can be related to them.

The final ingredient of the measurement component for this model is the collection of relevant educational context indicators which may be used to classify the school score distributions and thereby provide information useful in generating hypotheses about the antecedents of student performance. Examples of these indicators are socio-economic status and teacher verbal ability (see Coleman et al).

Data Collection

Preparing for the actual collection of data for this model requires a decision about the reference population. It may be a population of schools or individual students. If the school is the unit for analysis it is frequently easy to secure a sampling frame because most state departments maintain a list of operating schools in their states. The problem of obtaining

adequate representation within the schools however, is not so easily solved. Assuming a high degree of variability within a school, rather large student samples are required to provide a school score with a reasonable degree of precision. Therefore, it is recommended that for this model a sample of randomly selected schools be used to generate the state educational status data specified in the first objective. An example of this procedure follows.

Suppose that there are 3,000 elementary schools in the state and that the measure of interest has a standard deviation of fifteen. In this situation, a sample of 200 elementary schools will provide an estimate of the state average which will deviate from the actual average no more than two score points with 95% confidence. If information is available which is known to relate to the student characteristic of interest, stratification of the sample will result in greater precision, and a reduction of sample size becomes possible (see Kish).

It is possible, however, to select a sample which allows analysis of individual student data although the complexity and therefore the risk of administration error increases.

The design requires that the probability of selecting any one student remain equal even though we do not know the names of all students in the state. An example of this procedure is furnished in Appendix A.

The school sample described above does not permit the analysis of any individual student output data unless some generally untenable assumptions are made about the random assignment of students to schools. Neither does such a sample permit estimates to be made for individual school districts. Because of the anticipated substantial variance among schools within districts, it would probaby be necessary to sample most or all of the schools in each district in order to obtain suitable district estimates. As in the case of data collection at the state level, the model at the district level would not

necessarily require the testing of all students within the selected schools. However, such all-inclusive data collection might be more feasible than within-school sampling of individuals.

District estimates would, of course, enable districts to compare themselves to other districts "like them" as defined by a wide variety of situational variables. Reports could be generated to facilitate these comparisons in terms of both district averages and district score distributions.[1]

Analysis Procedures

The objective of this first level model is to provide a description of the position of the schools in the state with reference to the educational objectives deemed important in that state. The data may therefore be analyzed according to straightforward descriptive information. Frequency distributions for the state, the measure of central tendency, probably expressed as a mean, and a measure of variability, such as the standard deviation can readily be produced. The second objective of the model however, is to direct attention toward the different levels of achievement of student groups as a way of highlighting both differential difficulty of learning and the areas in which hypothesis generation might be productive.

For these purposes, analysis of the data should minimally include distributions classified by the qualitative information collected to reflect the educational context. This information does not have to be quantifiable, although the idea of levels may be appropriate, in order to be useful in the analysis. A minimum four cell classification of distributions is recommended, so that possible interactions may be detected. For example, the data might look like this:

---

[1] These district assessment ideas were suggested by Donald Trismen, Educational Studies.

Attitude Toward School

	High Staff Training		Low Staff Training	
High SES	47.8 49.5 46.2 52.1 51.0 51.7 54.7	$\bar{X}$ = 50.2	44.6 47.4 42.6 46.8	$\bar{X}$ = 45.8
Low SES	50.4 47.8 51.2 48.8	$\bar{X}$ = 49.5	50.2 47.3 52.0 47.6 54.5 46.8 44.3	$\bar{X}$ = 48.6

An inspection of these data suggests that, in terms of attitude toward school, teacher training has a more significant association than socio-economic level. There are appropriate statistical techniques for determining the probability that the observed differences are actual rather than due to chance. Two possible procedures are two-way factorial analysis of variance or a Friedman two-way analysis for ranked data.

The analysis described here is useful in identifying interrelationships which should be examined further. The purpose of this additional examination is to discover what experiences children have which may be modified to produce desired changes in output, in this case improving attitude toward learning. Observation of schools located in high scoring and low scoring groups in the classification tables, should suggest productive ways of changing the learning situation.

The limitations of this analysis, however, include the possibility that less distinct relationships may not be revealed. Also, it becomes extremely complex to examine the effects of several conditions taken as a group.

If the data collected on both student output and the conditions of learning are quantifiable, and if there is reason to believe that relationships are fairly uniform across the range of scores from high to low, it is appropriate to use multiple correlation techniques for the analysis. These procedures allow more complex relationships to be considered and provide a method for examining the unique contribution of many variables in a systematic way. Partial and semi-partial correlation techniques are included in this classification (see Nunally).

The principal use of these correlation techniques, as in the case of the two-way classification analysis, is to identify variables which should be studied for possible influence upon the experiences students have, and therefore upon what they learn. Results obtained from these techniques do not suggest corrective action directly, but are the first part of a two phase process of educational change. The second phase requires alteration of learning conditions or of the arrangement of learning experiences which can then be evaluated by a subsequent assessment.

Reporting Results

Interpretation of statistical data, particularly that which requires qualification, such as test scores, is most effective in a context in which interaction between the receiver and the presenter is possible. Therefore, the ideal method of interpretation includes a personal interface between the concerned school personnel and a presenter who knows both the nature of the data and the method of analysis. If the whole process is to be cost effective, a discussion of implications consistent with the results and suggestions of alternative courses of action must be included.

In reporting the results of the state status assessment to legislators, state boards of education, the governor's staff and other decision makers, the personal interface is extremely important. If standardized tests or NAEP

exercises and procedures have been used, the results can be compared with regional and national data.

If district or school building data have been collected, more detailed methods of reporting results are needed. Several options for reporting results are provided in Appendix B.

INTER-DISTRICT COMPARISONS

In the foregoing discussion the primary focus is on an assessment program that views the state in its entirety (i.e. the sample is drawn from the state population for the grade level(s) of interest). Considerable interest is being directed toward the use of student performance data gathered in an assessment to compare school districts.

The following is a brief discussion of two considerations that should be taken into account when inter-district comparisons are to be made. One relates to the sample design and the data analysis. The other pertains to socio-economic and other condition variables - both situational and individual - and their criticality when inter-district comparisons are to be made.

Sample Design and Data Analysis. As was indicated earlier some generally untenable assumptions must be made if district estimates are made from a sample of students where the population frame is the state in its entirety. Should inter-district comparisons be desired, the population from which schools or students are drawn should be the individual district. Since most school districts either maintain master lists of students or can obtain a list with minimum difficulty, the less complex method of drawing a sample is to use a simple random sample of students. It is possible, but more difficult, however, to use a two step cluster sample technique in some districts (i.e. in step 1 select a sample of schools; in step 2 draw a sample of students from each of the previously selected schools), while using a simple random sample procedure in others. This combined approach might be feasible when exceptionally large school districts do not maintain or have available student lists. In large districts the two step cluster sample technique might be used while in smaller districts where

student lists are available, a simple random sample of students could be drawn. The major difficulties in using the two step cluster sample technique relate to bias in local scores caused by the possibility of greater student homogeneity (i.e. less variability) in an individual school.

Most states have a wide range in the number of students enrolled in their school districts. When drawing a student sample, larger districts may be collecting data on, for example, 10% of the population, while in smaller districts it may be necessary to gather data from the entire population. When aggregating district data upon which to make inferences statewide, arithmetic weighting of district data is used when computing estimates of parameters and their standard errors. The disproportionate allocation of students in individual districts is compensated for by using inverse weights in the statistics. That is, in the larger district where 10% of the student population is sampled and therefore under-represented, the data are weighted up, while in a smaller district where data are gathered on all of the students and therefore over-represented, the data are weighted down. The correct weight for a district can be determined by $\frac{Nj}{N}$ where Nj is the number of students in the population from the jth district and N is the number of students in the total population.

Criticality of Condition Variables. When district comparisons are desired, special attention should be directed toward those conditions of learning that may be associated with student performance.

Student background characteristics such as socio-economic status, attitudes and aspirations have been found to be associated with (not to be confused with caused by) student achievement. Furthermore, other school variables such as the quality of the instructional staff (e.g. staff training) and the availability of financial resources also have shown an association with student achievement.

These variables become critical when inter-district comparisons are to

be made. For example, comparing an inner-city school district serving children from economically deprived families with the affluent suburbs surrounding could be grossly misleading. The inner-city school district, when the environment of its students is considered, might be making a more substantial contribution to student performance than its more affluent neighboring school districts, when the environment of their students is taken into consideration. Other examples could be cited; however, the important point is that as nearly as is possible, school districts should be compared with those that have similar characteristics that would be difficult to change by educational policy decisions (e.g. socio-economic status).

In comparing similar districts, it is important that categorization of districts reflect those difficult to change variables that are associated with differences in output. Although the ultimate objective of assessment is to provide information which will enable decision makers to improve the educational performance of all children, it is naive to expect that such improvement will result immediately. Therefore, the condition variables should be considered in designing alternative school programs which show promise of improving student performance (e.g. greater utilization of prior student experiences). Statistical procedures for determining categories include expectancy tables, regression analysis, analysis of variance and analysis of covariance.

SUMMARY

This plan provides some suggestions and ideas for initiating a statewide educational assessment program.

Guiding principles for an assessment should be: 1) to specify and define educational goals in terms of measurable outcomes; 2) to involve various publics extensively; 3) to use measurement instruments having face and content validity; 4) to include noncognitive student behaviors; 5) to present the results in a form understandable by those outside the professional education community; and finally 6) to view assessment not as an end unto itself but as a means of providing useful information to decision makers.

The objectives of the initial statewide assessment program should be: 1) to collect student performance data that can provide a status report on the quality of education in those goal areas identified as having high priority; 2) to introduce the concept of assessment and its usefulness as a source of information for both decision makers and concerned parents or taxpayers; 3) to provide a starting point whereby those managing the statewide effort may gain useful experience in operating the program; 4) to develop a method of data analysis that can illustrate the variability of performance due to individual differences among students and to the social context in which they live.

The essential components of the initial statewide assessment should include such introductory activities as the provision for an advisory committee and for selecting its members, conducting meetings to inform interested citizens and school personnel of the nature, scope and methodology of the assessment program. Other essential components are data collection materials, data analysis procedures and reporting strategies.

In order to make inter-district comparisons, the sample must be representative of the individual districts rather than the entire state. Simple random sampling of students in the districts is recommended when feasible. The more complex two step cluster sample technique may be used in large districts as an alternative.

When comparisons are made, the condition variables must be considered in categorizing the districts or the comparisons will not have meaning. Expectancy tables, regression analysis, analysis of variance and analysis of covariance are statistical procedures thay may be used to determine categories.

REFERENCES

Coleman, J. S. et al, Equality of Educational Opportunity. Washington: U. S. Government Printing Office, 1966.

Kerlinger, F. N., Foundations of Behavioral Research. New York: Holt, Rinehart and Winston, Inc., 1964.

Kish, L., Survey Sampling. New York: J. Wiley, 1965.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. Taxonomy of Educational Objectives, Handbook II: Affective Domain. New York: David McKay Company, 1956.

Nunnally, J. C. Educational Measurement and Evaluation. New York: McGraw-Hill 1964.

APPENDIX A

Sample selection - individual student sampling

unit - equal probability

a. Arrange the schools in any order, with the number of students in the grade of interest specified for each school.

b. Assign sequential blocks of numbers to each school. The size of the block is determined by the number of students in the target grade. Inflate the block size by a specified percentage to allow for increases in enrollment. This percentage should be taken into account when determining sample size so that the desired sample size will remain after the shrinkage caused by empty sampling frames (i.e. numbers in the block not having a corresponding individual).

c. Using a suitable random selection procedure, (table of random numbers, computer program, etc.) determine the numbers which are to be included in the sample.

d. Assign selected numbers to each school block, converting them to the sequential number within the block (eg. in the school with block 2998-3092, 2998 becomes 1, 2999 becomes 2, 3024 becomes 27.)

e. Instruct school coordinators to arrange class lists in any order and assign the numbers 1 through N to the <u>total</u> group of students (not recycling through classes.)

f. <u>After</u> the assignments have been made, the instructions are to open the envelope containing the randomly selected numbers and administer the test to those students whose names on the school list correspond to these numbers.

APPENDIX B

OPTIONS FOR REPORTING RESULTS

Option One - Interface

Step 1

Process data and prepare written reports

A. Alternative formats

1. Expectancy tables based on previous year's performance.
2. Comparison with state norms.
3. Percentage of responses to each option of "key" items--those where content is face valid and which relate most highly to other items in the subscale, thereby reflecting well the concept measured.
4. Description of the distribution of children in terms of the kinds of problems they are successfully solving and the kinds which are presenting difficulty. If available, report response patterns for high scorers, middle scorers, and low scorers.

Step 2

Assemble and train a corps of data interpreters.

A. Sources of personnel

1. State department personnel
2. College and university staff, graduate students, interns, etc.
3. School personnel
4. Intermediate unit office staff

No single source can provide enough personnel to comprise the necessary teams, but a combination may make possible sufficient numbers to provide one

interpretive visit to each district.

B. Training

1. Sessions concerning test development, item to goal correspondence, reliability, group statistics versus individual statistics, and presentation of comparative data classified by school variables. Methodology includes presentations using actual state data.
2. Sessions exploring implications of results and identifying strategies for the development of alternative action plans where need is indicated. These sessions should include specialists in the areas assessed. The objective of these sessions is to stimulate action through provision of suggested approaches or avenue of exploration based on assessment results.

Step 3

A. Schedule presentation sessions with each school district. It is recommended that a work group comprising administrative staff, teachers, parent advisory representatives, at least one board member, and other interested people be included.

Step 4

Organize a consortium of educators across the state who can serve as a resource to the school personnel who will be working on altered approaches to learning and personal development for their students.

The resource requirement for a program of this sort is dependent in part upon voluntary professional commitments of the state's educators. It also requires allocation of funds for personnel time on the part of many institutions in addition to the Department of Education. If properly approached, there is enough professional commitment among professional educators to achieve some degree of this resource allocation.

For a state with 500 districts the requirements of the interpretation team are approximately as follows. In a one month period the 500 districts would be visited on the basis of 1-1/2 days per district. Allowing some surplus for contingencies, 78 persons working 11 days each will be required. Based on a per diem expense of $25 plus an average transportation cost of ten cents per mile for 150 miles, the total cost will be less than $30,000.

Such an approach to the problem of reporting and utilizing assessment data has the best chance of not only assuring positive follow up but also of reducing hostility in both the schools and their communities.

Option Two - Regional Interface Model

It is recognized, however, that resource constraints may not allow so comprehensive a commitment. Therefore, a less optimistic alternative can be conceived.

This model would follow the same basic steps outlined in Option One, but rather than providing individual district data presentations as outlined in Step 3, it would utilize a series of regional meetings in which about two representatives from each district would participate. The format suggested for the regional meetings is the model used for the training sessions designed for the data interpreters (Step 2 of Option One). The corps of data interpreters in this case would be smaller since 30 districts can be accommodated in each regional meeting requiring only four data interpreters for each session. The person days required are thus reduced to about 120. If we assume that local districts will pay transportation and housing costs of their representatives, and also assume the professional investment of the state's educators, the cost for this kind of program would be about $6,000. It must be recognized that the technical experts used for training data

interpreters will not be available for the 17 regional meetings and that their contribution would be filtered through two "student" levels, the data interpreter and the school district regional meeting attendees. However, the concentrated thinking about the problems and the utilization of assessment results that such a program would encourage would still stand a fair chance of having an impact on the education in a state.

Option Three - Lecture-Discussion Model

In order of desirability, a third level of data presentation might be to conduct five regional meetings with an average of 200 participants comprising two representatives from each of one hundred districts. Real but non-identifiable state data could be presented in a visual lecture-discussion mode in an auditorium setting with questions limited to a small sub-group of school representatives. These interrogators could collect their questions from discussion sections prior to the question and answer period. Such meetings could be handled by State Department personnel augmented by some representation from the corps of experts available within the state. Following these regional meetings the assessment data would be released to each local school district--preferably scheduled to allow time for district staff consideration before it becomes public.

Option Four - Mail-out Kits

The least desirable, least expensive and not only least useful, but possibly useless, presentation is a mail-out kit of charts, lists and printed discussions which would be sent to the superintendent in each district. To optimize its very slim chance for utility and positive impact, the mail-out kit should include the following components.

First, the data should reflect the maximum amount of knowledge on the part of the State Department of Education about the school district to which it is directed. Second, it should be personalized in a form which describes what the students are like who represent several points on the distribution of data for the school.

For example, a student in the upper quarter of the score range on the reading scale could be described in terms of skill content of the question to which he responds correctly. The actual pattern of responses should be included. The report could read "Students in the highest quarter of the score range tend to answer correctly items which require inference two out of three times, and items which require locating factual detail four out of five times. There are 178 (20%) students in this range from your school. For comparison, students from other schools of your community type (Type 3) respond correctly to inference items three out of four times and to factual detail questions also three out of four times. Seventeen percent of all community Type 3 students score in the highest quarter."

A similar discussion of middle and low scorers should be provided. Backup material describing community type and resource availability should also be included. In addition, any of the written alternate forms of presentation suggested in Option One as material for the more desirable personal interpretation could be provided. This published form of presentation could also be used as a supplement to any of the options.

In summary, the most productive interpretive format for State Assessment data is conceived to be an individually tailored personal presentation to school district personnel. The least productive interpretation is a written report. A State Department of Education is urged to locate the resources to do personalized data interpretation rather than a less involved and therefore less expensive mail-out.